PREFACE

Link prediction is a task that involves network structure evaluation and identifying missing and future links in social networks. In dynamic networks, each edge is associated with a temporal component, and link prediction in such networks is the task of identifying edge existence probability based on previously observed network dynamics. This research is crucial to comprehending network evolution and its effects on individual node behavior with respect to time. Link prediction algorithms can be applied to extract missing data, identify erroneous interactions, evaluate the factors governing network evolution, and more. Node pair similarity-based link prediction methods use different structural or topological information to calculate possible edges' chance. In the snapshot-based paradigm for link prediction on dynamic graphs, feature sets created from combining different similarity-based methods are used to track the temporal changes in edges. These features are then used to make predictions for non-existing edges. Machine learning algorithms are used on these feature sets to build prediction models. Compared to the individual similarity-based approaches, machine learning techniques have been used to improve prediction accuracy as they are more helpful in modeling changes in patterns over time. This thesis aims to enhance link prediction on dynamic networks using different feature sets and machine learning algorithms.

The following four fundamental aspects of link prediction are explored in this thesis since they are used to build the feature sets for machine learning classifiers:

1) incorporating local, global, and quasi-local similarity indices for link prediction, 2) merging snapshot-based features with snapshot-impartial path-based and cost-based features, 3) feature selection to identify best feature sets based on similarity indices, 4) exploring community detection enhanced link prediction features, and 5) studying the effect of quantum kernel-inspired transformation on link prediction in dynamic networks.

One of the well-known categories of link prediction methods is the similarity-based method, which uses a node pair similarity score for edge estimation. The three widely used categories of similarity-based indices are Local (L), Global (G), and Quasi-local (Q) indices. All three types of these methods can be combined to create a rich feature set that would be more useful for link prediction than its components individually. The proposed LGQ framework explores the relative effect of combining these categories of link prediction features on different machine learning algorithms.

Further exploring the role of snapshot-impartial paths, Path Weight Aggregation Feature (PWAF) for Link Prediction in Dynamic Networks is proposed. Different topological aspects of the networks (Local, Global, and Quasi-local), as well as Clustering Coefficient based features, are considered for feature generation, in addition to the suggested Path Weight-Based Aggregation Feature (PWAF). Some of these features also take advantage of longer paths between nodes for a more thorough similarity estimation.

A feature-based solution that considers individual snapshots and the overall network throughout the full-time span to answer the link prediction problem is explored. A novel feature called Cost-based feature for link prediction (CFLP) for estimating edge behavior throughout the entire network is proposed, which uses a reward and penalty structure to summarize node activity across the whole network (snapshot-impartial), is presented. The feature set also uses similarity indices, classified into four major categories: local similarity, global similarity, quasi-local similarity, and clustering coefficient-based similarity, to measure edge activity change in individual snapshots. Different feature selection methods are used to correctly quantify the relative effect of features among themselves and the overall link prediction problem.

The highly connected groups of nodes within the network are called communities, which have an important role in understanding and uncovering various functional properties of the system. A community information-based feature estimation and link prediction (COMMLP) method applied to dynamic graphs in a per snapshot feature estimation-based setting is proposed to take advantage of the group behavior of nodes. First, a link prediction framework is presented to predict missing links using parameterized influence regions of nodes and their contribution to community partitions. Then, a unique feature set is generated using local, global, and quasi-local similarity-based and community information-based features. This feature set is further optimized using scoring-based feature selection methods to select only the most relevant features.

Finally, the solution to the issue of link prediction in dynamic networks using supervised learning and the Projected Quantum Kernel (PQK) is addressed. It is conceivable to

expect that quantum computers may outperform classical computers in machine learning tasks since quantum systems display aberrant behavior that conventional systems are thought to be incapable of creating. The QML approaches use the huge dimensionality of quantum Hilbert space to get an optimized solution by modeling the feature space of a classification problem with a quantum state. The projected quantum kernel-based link prediction (PQKLP) approach is presented to address the link prediction problem using both local and global information. Using PQK, features are transformed from the popularly used snapshot-based feature set form into quantum space such that the effectiveness of machine learning-based classification can be improved.